DADAED – DOUBLE ANOMALY DETECTOR WITH AEDIFF

Technical Report

Jan Tožička, Marek Bezůšek, Karel Durkota, Michal Linda

NeuronSW SE
{jan.tozicka, marek.bezusek, karel.durkota, michal.linda}@neuronsw.com

ABSTRACT

This report describes our submissions to the DCASE 2022 challenge Task 2 ”Unsupervised Detection of Anomalous Sounds for Machine Condition Monitoring under Domain Shifted Conditions.” Acoustic-based machine condition monitoring is a challenging task with a very unbalanced training dataset. Moreover, due to domain-shift, testing data may come from a different distribution than the training data, which makes the task even more difficult.

In this submission, we propose two novel extensions of anomaly detection based on the reconstruction of auto-encoder (AE) network. The first approach uses the raw difference between AE input and its reconstructed output (instead of typical reconstruction error based anomaly detectors). The second approach extends the first approach with an additional anomaly score of auto-encoder's latent vectors. The combination of these two anomaly scores is then used to determine the final anomaly score.

Index Terms— Predictive Maintenance, Anomaly Detection, Auto-encoder, OpenL3

1. INTRODUCTION

Machine condition monitoring is an essential component of predictive maintenance. It allows to schedule maintenance work to fix machine problems in the earliest stages and thus reducing maintenance costs and preventing consequential damages. Acoustic emission monitoring can be used for machine condition analysis and prognosis. ISO 22096[1] suggests that the nature of acoustic emissions can be used even without an understanding of the operating mechanics of the monitored machine.

The recent progress in AI allows us to create an automatic machine condition monitoring system. However, most of the AI methods require a huge amount of well-labeled examples, which makes them difficult to apply for machine condition monitoring tasks, where, it is often impossible to collect all failures. The reason for the lack of the data is that each machine failure may sound slightly different, and in order to have many sounds of broken machines, we need to have many broken (often very expensive) machines. In practice, it is exceptional to get even a few examples of a failed state.

Additionally, due to the machine's different operating statuses or just the machine's wear-and-tear, the emitted nominal sounds may shift, which is called domain-shift. This introduces an additional complexity to the problem, because training data may not even contain some specific machine statuses (e.g. speed, configuration). Therefore, AI approaches must be capable of generalizing from the available data. The Task 2 in DCASE2022 challenge is set up very realistically, meaning that (i) only nominal data are available for training and (ii) contains domain-shift.

Among several options, we chose to use reconstruction-based anomaly detector with auto-encoders (AEs). AEs learn to reconstruct the input features using smaller latent space. For AEs to be able to reconstruct well the input, it must reasonably encode its inputs into the latent space, which means extracting important information about the sounds. We use reconstruction together with the intermediate latent space as inputs to classical anomaly detectors (KNN, LOF) to obtain anomaly score, described in details in the next sections.

2. OUR APPROACHES

We used pretrained OpenL3[1, 2] neural network to convert audio embedding space, over which we continue with trained our auto-encoders, as described in Figure 1. In detail, each audio sample is converted with OpenL3 into 26 matrices m_i of shape 96×6144. We compute mean μ and σ of matrices $m_1, \dots m_{26}$, and concatenate them $x = \langle \mu_i, \sigma_i \rangle$ into a single vector for an input into the auto-encoder. Auto-encoder parametrized with θ learns to reconstruct the input x in its output $x' = \theta(x)$ using a smaller intermediate latent space (32 dimensional in our case). Formally, minimizes

$$\arg\min_{\theta} ||X - \theta(X)||^2$$

In [3], Lehtinen et al. describe that every auto-encoder actually works as denoising auto-encoder. Thus the reconstruction error is not influenced only by the anomalies but also by the noise, which is significant in DCASE domains. Therefore, we propose to use the difference $\delta = x - \theta(x)$ as new feature vector, on which we want to detect anomalies. δ is processed with PCA to reduce its dimensionality to 30. Afterwards, Local Outlier Factor (LOF), or KNN, is used to get the actual anomaly score. Average of anomaly scores of all 26 chunks for single sample produces final anomaly score in our first approach, called AEDiff.

Our second approach, DADAED extendes AEDiff by adding second anomaly detector (LOF, or KNN), which detects anomalies on the intermediate latent vector. The two anomaly scores (one from AEDiff and one from the new anomaly detector) are then probabilistically combined with disjoint OR operator $P(A \text{ or } B) = 1 - (1 - P(A))(1 - P(B))$.

Our approaches use one model for all domains (except valve domain) to prevent overfitting. We do not use any information about domain shift, which increases applicability of our approach (such information is not always available).

[1] https://www.sis.se/api/document/preview/908883/

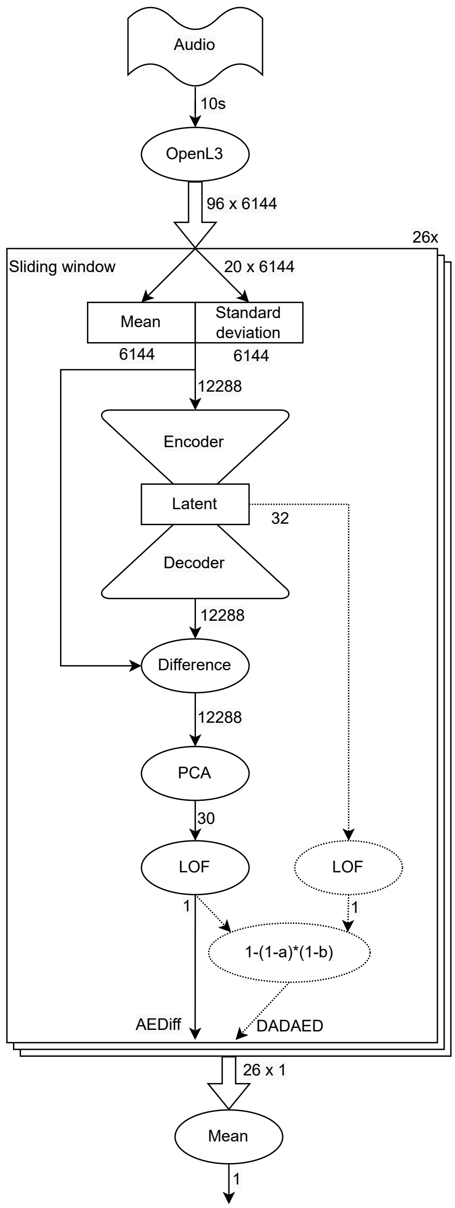

Figure 1: Scheme of the anomaly detection model. The dotted part shows how the AEDiff model is extended by a second anomaly detector – this extension is called DADAED (Double Anomaly Detector AEDiff). Numbers show the dimensions of vectors.

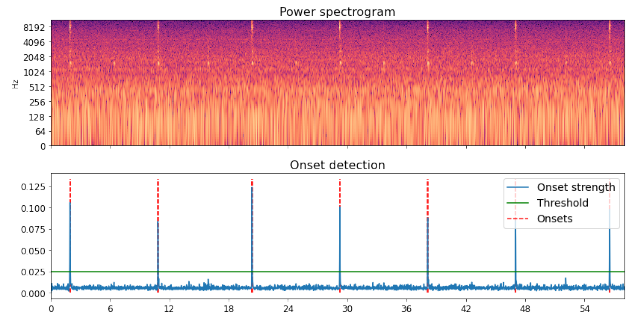

Figure 2: Tick onset detection in valve sounds. Comparison of power spectrogram calculated from one sample with onset strength envelope.

3. MODEL FOR VALVES

The valve sounds are significantly different from the sounds of other machines. Most of the valve sounds are noise with only a few tenths of a second containing the relevant sound where the valve is opening or closing. Thus, all above described models failed to detect anomalies with AUC below 55 %.

Therefore, we designed a special model for valves, which calculates the energy of the opening and closing valve sounds. During exploratory data analysis, we found a loss of energy at around 4 – 8 kHz in the valve ticks in anomalous samples. The hypothesis is that this change in sound might correspond to the simulated anomaly condition – a small piece of paper caught in one or more of the valves.

The algorithm for detecting and isolating the valve's ticking sounds consists of the following steps:

- calculating the Short-time Fourier transform,
- high-pass filtering with a cut-off frequency of 4 kHz,
- ticks onset detection,
- selecting the entire sound of each tick with a fixed window (0.25 second).

For the onset detection, we used algorithms from the librosa package – spectral flux onset strength envelope and peak finding in an onset strength envelope. A threshold was set to find only relevant peaks in the strength envelope (see Figure 2).

The energy of each windowed tick was then calculated and averaged for the entire 10-second long sample. The negative energy was used as the anomaly score. This model outperformed all other models on the development dataset for the Valve machine with AUC: 86.4 % and pAUC: 76.2 %. Thus, it is used in all submissions.

4. SUBMISSIONS

We have submitted four variants of our approaches. In all variants we used OpenL3 with parameters as follows:

- Input Representation: *linear*
- Content Type: *music*
- Embedding Size: *6144*
- Hop Size: *0.1*

and our auto-encoder with architecture as follows:

	Submis. 1		Submis. 2		Submis. 3	
problem	AUC	pAUC	AUC	pAUC	AUC	pAUC
bearing	**67.4**	**61.0**	65.8	54.7	64.0	53.5
fan	62.0	60.4	61.6	58.9	**67.1**	**60.6**
gearbox	**79.8**	**64.8**	77.7	66.3	72.1	59.6
slider	77.0	62.6	**81.0**	**68.4**	77.6	58.7
ToyCar	87.8	67.8	**89.4**	**71.1**	86.3	65.4
ToyTrain	**69.0**	**63.8**	67.7	61.8	68.6	60.5
Average	74.5	64.9	74.3	64.6	73.7	61.4

Table 1: Harmonic means of AUCs and pAUCs on development dataset. Sumbission 4 is composed of highlighted models for specific machines. It scored AUC: 76.2 %, pAUC: 66.2 %.

- Input (12288)
- Dense (128) + ReLU
- Dense (32) + ReLU
- Dense (128) + ReLU
- Output (12288)

The four submissions are as follows:

Submission 1 - uses OpenL3 prepossessing, our AEDiff training and LOF final anomaly detector.

Submission 2 - uses OpenL3 prepossessing, our DADAED training and LOF final anomaly detector.

Submission 3 - uses OpenL3 prepossessing, our AEDiff training and KNN final anomaly detector.

Submission 4 - combines predictions from all the above mentioned approaches – for each machine type the predictions which performed the best on the development dataset has been chosen. Namely:

- ToyTrain, Bearing, Gearbox - predictions from Submission 1,
- ToyCar, Slider - predictions from Submission 2,
- Fan - predictions from Submission 3,
- Valve - our custom model.

This Submission achieved the overall scores AUC: 76.2 %, pAUC: 66.2 %.

5. RESULTS

To evaluate all three approaches, we used Development data of DCASE2022 Task-2 Challenge[4, 5, 6]. In Table 1 we summarize the AUCs and pAUCs for $p = 0.1$ of all four submissions (for detailed results see Appendix). We can see that all submissions have similar results with small differences in some domains (the best scores are highlighted in the table). The Submission 4 combines the remaining submissions by choosing the best performing method for each machine type.

6. CONCLUSION

We have proposed and evaluated a novel anomaly detector based on auto-encoder's reconstruction difference. One approach uses solely the reconstruction difference, while second approach uses additional data from latent space of auto-encoder. The submission combining these two approaches results with the highest average score overall of AUC: 72.6 %, and pAUC: 66.2 %, outperforming the baseline AutoEncoder solution.

7. ACKNOWLEDGMENT

We want to thank NeuronSW SE company, for the support and for providing computational resources to train and evaluate the algorithms.

8. REFERENCES

[1] J. Cramer, H.-H. Wu, J. Salamon, and J. Bello, “Look, listen, and learn more: Design choices for deep audio embeddings,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 05 2019, pp. 3852–3856.

[2] R. Arandjelović and A. Zisserman, “Look, listen and learn,” in *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[3] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, “Noise2noise: Learning image restoration without clean data,” *CoRR*, vol. abs/1803.04189, 2018. [Online]. Available: http://arxiv.org/abs/1803.04189

[4] K. Dohi, K. Imoto, N. Harada, D. Niizumi, Y. Koizumi, T. Nishida, H. Purohit, T. Endo, M. Yamamoto, and Y. Kawaguchi, “Description and discussion on DCASE 2022 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques,” *In arXiv e-prints: 2206.05876*, 2022.

[5] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, M. Yasuda, and S. Saito, “ToyADMOS2: Another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions,” in *Proceedings of the 6th Detection and Classification of Acoustic Scenes and Events 2021 Workshop (DCASE2021)*, Barcelona, Spain, November 2021, pp. 1–5.

[6] K. Dohi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, Y. Nikaido, and Y. Kawaguchi, “MIMII DG: Sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task,” *In arXiv e-prints: 2205.13879*, 2022.

Appendix

Detailed results for each class is shown in Table 2.

	Submis. 1		Submis. 2		Submis. 3		Submis. 4	
problem	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC
fan-0 S	58.12	60.00	48.32	54.74	82.24	68.84	82.24	68.84
fan-0 T	71.16	62.32	71.44	60.63	71.12	59.58	71.12	59.58
fan-1 S	61.40	57.05	62.64	54.11	59.00	53.68	59.00	53.68
fan-1 T	47.04	51.58	55.32	52.21	57.12	52.00	57.12	52.00
fan-2 S	76.60	78.95	76.40	75.16	74.64	77.26	74.64	77.26
fan-2 T	66.72	58.74	64.04	61.68	65.32	59.16	65.32	59.16
gearbox-0 S	92.88	86.53	87.08	78.95	84.20	72.21	92.88	86.53
gearbox-0 T	75.24	64.42	84.64	74.32	69.96	56.42	75.24	64.42
gearbox-1 S	81.96	57.68	78.40	62.74	74.84	51.16	81.96	57.68
gearbox-1 T	73.72	56.00	62.24	53.68	61.72	51.58	73.72	56.00
gearbox-2 S	79.00	68.42	85.08	77.89	75.24	73.68	79.00	68.42
gearbox-2 T	78.52	63.79	72.44	58.95	70.52	60.00	78.52	63.79
bearing-0 S	82.68	69.26	67.36	52.03	64.68	54.74	82.68	69.26
bearing-0 T	77.44	59.16	69.52	53.16	69.68	51.01	77.44	59.16
bearing-1 S	43.84	47.58	53.76	51.32	53.48	48.00	43.84	47.58
bearing-1 T	82.64	56.84	84.92	68.21	85.08	65.26	82.64	56.84
bearing-2 S	67.68	70.11	61.50	51.11	62.34	53.01	67.68	70.11
bearing-2 T	69.68	70.32	65.20	55.59	57.42	52.11	69.68	70.32
slider-0 S	83.88	72.21	88.32	83.37	87.76	70.11	88.32	83.37
slider-0 T	79.80	57.26	85.60	64.63	80.68	49.47	85.60	64.63
slider-1 S	85.76	62.32	91.76	73.68	83.68	58.95	91.76	73.68
slider-1 T	81.32	60.84	80.96	66.74	77.32	58.74	80.96	66.74
slider-2 S	76.84	65.68	78.40	69.89	78.40	58.95	78.40	69.89
slider-2 T	60.76	59.58	66.32	57.68	62.84	59.58	66.32	57.68
valve-0 S	99.92	99.58	99.92	99.58	99.92	99.58	99.92	99.58
valve-0 T	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
valve-1 S	63.80	48.21	63.80	48.21	63.80	48.21	63.80	48.21
valve-1 T	87.04	71.58	87.04	71.58	87.04	71.58	87.04	71.58
valve-2 S	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
valve-2 T	81.60	71.58	81.60	71.58	81.60	71.58	81.60	71.58
ToyCar-0 S	87.40	64.21	84.68	54.11	86.36	59.37	84.68	54.11
ToyCar-0 T	81.68	56.42	87.60	67.58	83.44	60.21	87.60	67.58
ToyCar-1 S	82.04	61.05	88.64	78.74	84.52	65.68	88.64	78.74
ToyCar-1 T	86.00	65.47	90.12	73.47	78.96	60.63	90.12	73.47
ToyCar-2 S	99.56	97.68	100.00	100.00	99.96	99.79	100.00	100.00
ToyCar-2 T	92.76	75.16	87.08	67.79	87.08	60.42	87.08	67.79
ToyTrain-0 S	63.80	60.00	66.04	62.11	62.80	55.16	63.80	60.00
ToyTrain-0 T	53.24	50.53	49.00	47.79	54.20	48.21	53.24	50.53
ToyTrain-1 S	85.20	70.74	83.88	71.37	80.88	65.05	85.20	70.74
ToyTrain-1 T	57.32	57.26	56.40	56.63	57.64	56.00	57.32	57.26
ToyTrain-2 S	96.12	84.21	95.76	84.84	93.32	79.79	96.12	84.21
ToyTrain-2 T	77.48	70.95	76.76	60.00	79.20	69.05	77.48	70.95
Average	74.51	64.88	74.26	64.61	73.68	61.44	76.20	66.16

Table 2: Detailed results for each section, source(S) and target (T).